



UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
DEPARTMAN ZA MATEMATIKU I
INFORMATIKU



REPORT

Praktični zadatak iz predmeta Socijalne mreže

dr Miloš Savić, profesor

Dušica Knežević, asistent

Filip Vasić 50/19 IT

Нови Сад, 2022.

Sadržaj

Uvod	5
O kodu	5
Testiranje.....	7
Hard coded mreže.....	7
Random mreže.....	15
Erdos-Renyi model.....	15
Gilbert model.....	18
Watts-Strogatz Model.....	20
Barabasi-Albert model	22
Realne mreže	24
Slashdot mreža	24
Epinions mreža	27
WikiRfA mreža	29
Zaključak.....	32

Uvod

Praktični projekat iz predmeta Socijalne mreže predstavlja implementaciju funkcionalnosti i testova u cilju ispitivanja klasterabilnosti označenih neusmerenih mreža. Za mrežu kažemo da ima ovu osobinu tj. da je klasterabilna ako njene čvorove možemo podeliti u k koalicija tako da unutar koalicija nema negativnih grana, tj. da su čvorovi unutar koalicija međusobno povezani isključivo pozitivnim vezama, dok su veze između različitih koalicija negativne.

Praktični projekat obuhvata implementaciju algoritma za proveru da li neka mreža klasterabilna i ako nije daje mogućnost vraćanja linkova koji kvare osobinu klasterabilnosti tj. onih linkova koja je potrebno izbrisati kako bi ispitivanja mreža postala klasterabilna. Takođe implementacija omogućava isporučivanje svih klastera u mreži korisniku kao i posebno onih klastera koji jesu ili nisu koalicije, takođe moguće je isporučivanje klastera i kao mreže gde jedan klaster predstavlja jedan čvor koji u sebi nosi informaciju koliko čvorova taj klaster zapravo sadrži u sebi.

U okviru projekta biće izvršeni testovi na tri kategorije mreža. Testovi će se prvo raditi na malim „hardkodovanim“ mrežama za koje znamo jesu li ili nisu klasterabilne, zatim će se testovi vršiti na nasumično generisanim poznatim mrežama rađenim na predavanjima, dok će na kraju kroz testove proći i realne označene mreže kakve su Epinions, Slashdot i WikiRfa mreže.

Na kraju testiranja mreže će biti sačuvane u GML formatu kako bi se kasnije mogle vizuelizovati u alatu Gephi, koji je rađen na vežbama tokom kursa.

Implementacija samih funkcionalnosti i testova biće odrađena u programskom jeziku Java uz korišćenje *Java Universal Graph Framework* (JUNG) biblioteke za rad sa grafovima. Funkcionalnosti i testovi biće implementirani sa gledišta nadogradnje JUNG biblioteke tj. kao generičke Java klase.

O kodu

Kod implementacije je dostupan na mom ličnom github nalogu tj. na sledećem linku: https://github.com/fvasicc/SOCNETprakticni_zadatak.

Implementacija je rađena sa gledišta proširenja JUNG biblioteke tako da su sve klase u programu generičke i primaju parametre proizvoljnog tipa. U paketu `clusterability` nalazi se generička klasa `ComponentClustererBFS` koja implementira dva interfejsa `ComponentClustererU<V, E>` i `MarkedGraphMetricsU<V, E>` koji propisuju metode potrebne za određivanje komponenti ulazne mreže odnosno metode za određivanje da li je graf klasterabilan, koji klasteri su koalicije, koje grane kvare osobinu klasterabilnosti i sl.

Paket `metrics.clustering` sadrži klasu `ClusteringCoefficient` koja je takodđe generička i nudi metode za određivanje prosečnog koeficijenta klasterisanja, pronalaženje čvora sa najvećim koeficijentom i sl.

Svi interfejsi se nalaze u paketu `interfaces`.

U paketu `metrics.smallworld` se nalazi klasa `SmallWorldCoefficient` koja nudi metode za rad sa metrikama separacije tj. sa metrikama malog sveta. Nudi mogućnost računanja *small-world koeficijenta* i efikasnosti mreže

U paketu `metrics.centrality` nalazi se klasa `CentralityMetrics` koja nudi neke metode za rad sa metrikama centralnosti, npr. vraća čvor sa najvećom jednom od tri centralnosti ili vraća n čvorova sa najvećom jednom od tri centralnosti *betweenness*, *closeness* ili *eigenvector*. Kod *betweenness* centralnosti možemo birati da li želimo originalne vrednosti ili skalirane na interval $[0,1]$.

U paketu `model` nalaze se klase potrebne za predstavljanje čvorova i veza u grafu. Za označene veze u grafu koristi se klasa `MarkedEdge` iz paketa `model.edge` koja implementira generički interfejs `Marked` koji omogućava predstavljanje pozitivnih i negativnih veza u grafu. Za predstavljanje znaka veze koristi se enum `Mark`, iz istog paketa, koji daje dve mogućnosti, veza je ili pozitivna ili negativna.

U paketu `model.node` se nalazi klasa `GCNode` koja se koristi za predstavljanje čvorova u mreži komponenti. Ova klasa pored osnovnih polja ima i polje *nodes* koje predstavlja koliko se čvorova nalazi u originalnoj komponenti grafa koju predstavlja ovaj čvor.

U paketu `networks` nalaze se klase koje omogućavanje kreiranje proizvoljnih mreža na osnovu modela koji su rađeni na predavanjima. Sve klase su genetičke i sve implementiraju interfejs `RandomGraph` koji propisuje implementaciju samo jednog metoda `void getGraph(UndirectedSparseGraph<V, E> targetGraph)`. Metod ne vraća vrednost a kao parametar prima mrežu u kojoj će kreirati mrežu. Sve klase u konstruktoru primaju objekte tipa `Supplier` koji služe za kreiranje čvorova i linkova.

U paketu `tests` nalaze se klase koje su potrebne za testiranje kao i klase za čitanje i pisanje grafova iz odnosno u fajl. Klasa `NetworkReader` nudi mogućnost čitanja realnih mreža na kojima će graf biti testiran dok klasa `NetworkWriter` nudi mogućnost ispisa grafova u GML format.

Paket `tests.hardcoded` sadrži klasu koja služi za testiranje funkcionalnosti na malim mrežama koje su napravljene tako da znamo da li jesu ili nisu klasterabilne. Paket `tests.random_networks` sadrži klase koje se oslanjaju na prethodno navedene klase za kreiranje proizvoljnih grafova, a u paketu `tests.real_networks` nalaze se klase za testiranje na realnim mrežama.

Klasa `PrettyPrint` služi samo za ispis menija i obradu korisničkog unosa, odnosno da vrati odgovor korisniku za upit koji je izabrao.

U *res* folderu se nalaze svi fajlovi koji su korišćeni za testiranje kao i GML fajlovi koji su generisani tokom testiranja.

Testiranje

U ovom delu će biti prikazane funkcionalnosti programa kroz testiranje na najpre veštačkim mrežama koje su hardkodovane, zatim na random generisanim mrežama na osnovu modela sa predavanja, a na kraju i na realnim mrežama. Osim testova biće dati vizuelni prikazi mreža u alatu Gephi.

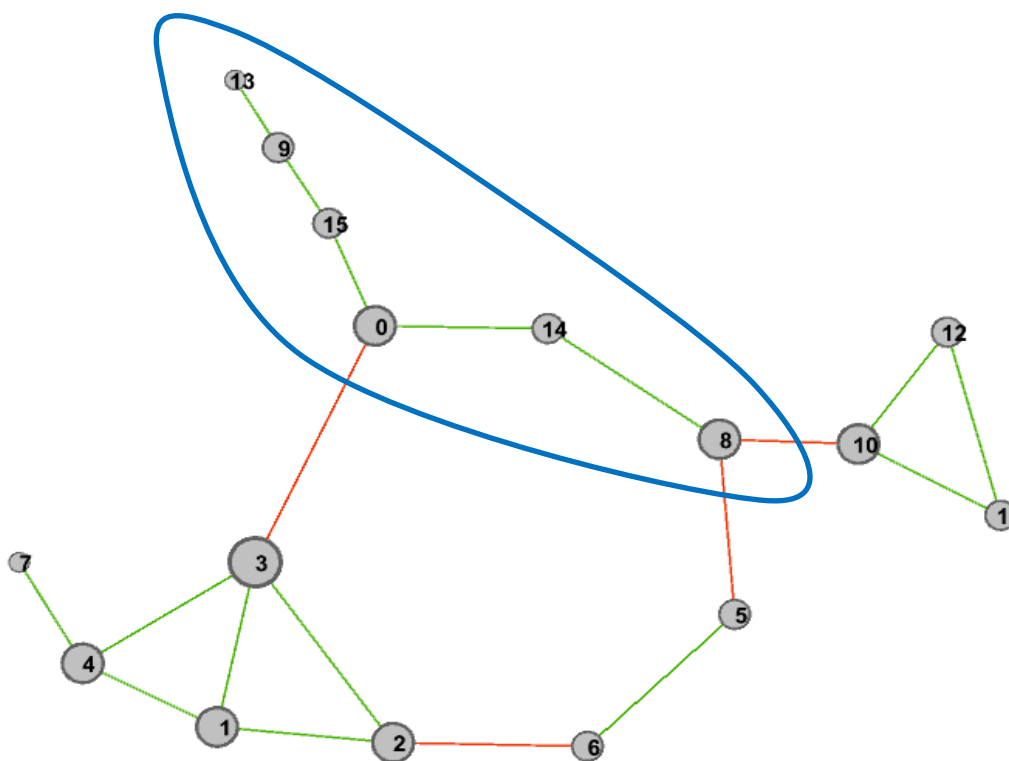
Hard coded mreže

U nastavku će biti prikazane analize veštačkih mreža koje se nalaze u fajlovima *clustered.txt1*, *clustered2.txt*, *nonclustered1.txt*, *nonclustered2.txt*, *nonclustered3.txt*, *nonclustered4.txt*, gde su prve dve mreže klasterabilne a ostale nisu.

Sve mreže su kreirane nasumično tako da je vođeno računa o tome da prve dve mreže imaju osobinu klasterabilnosti a da ostale četiri nemaju. Neklasterabilne mreže su samo modifikacije klasterabilnih takve da su dodati neki novi linkovi ili su izmenjeni neki znakovi veza da bi se izgubila osobina klasterabilnosti.

clustered1.txt

Klasterabilna	Ukupno klastera	Klasteri koalicije	Klasteri koji nisu koalicije	Broj čvorova	Broj pozitivnih linkova	Broj negativnih linkova
Da	4	4	0	16	15	4



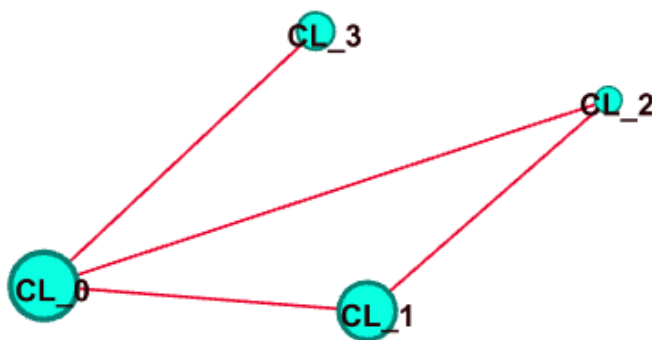
Slika 1

Na slici 1 prikazano je kako izgleda graf koji se analizira. Na slici je veličina čvora direktno proporcionalna njegovom stepenu a veze između čvorova su predstavljene odgovarajućim bojama, crveno negativni linkovi, zeleno pozitivni linkovi.

Mreža je klasterabilna pa nema problema sa negativnim vezama.

Klasterisanjem mreže dolazi se do 4 klastera od kojih su svi koalicije. Kako izgleda mreža klastera prikazano je na slici 2. Veličina čvorova na grafu je direktno proporcionalna broju čvorova unutar klastera.

Gigantska komponenta u grafu sadrži 6 čvorova i na slici 2 je označena sa CL_0, dok je na slici 1 uokvirena plavom bojom.

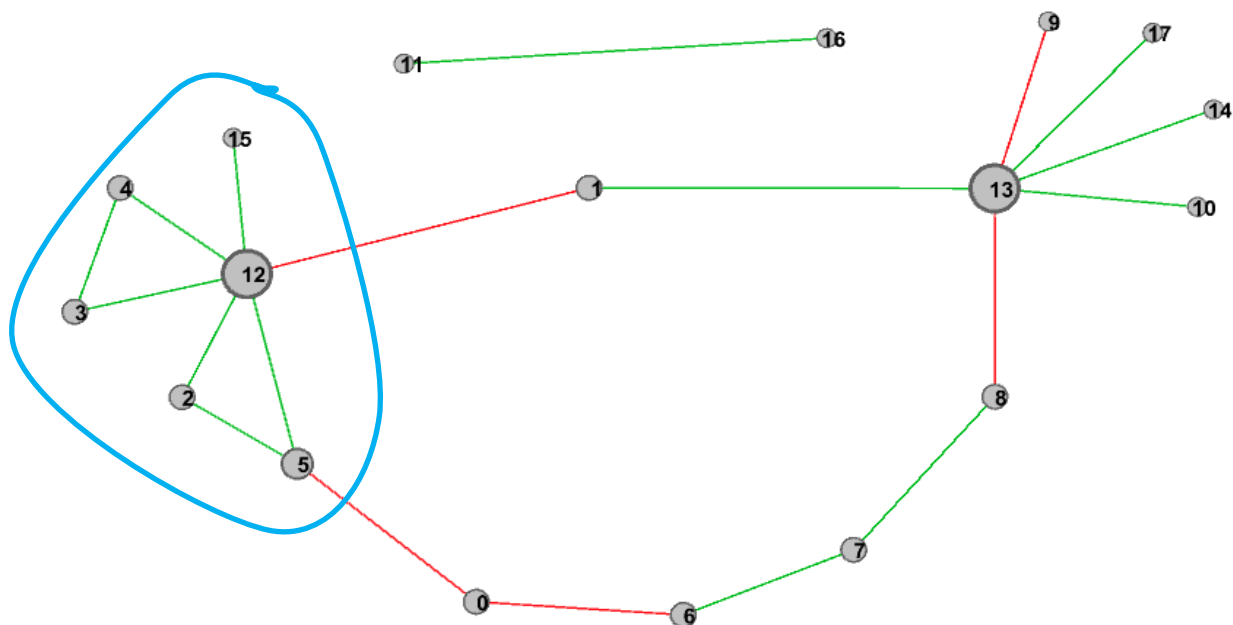


Slika 2

Treba samo napomenuti da su sve veze na grafu koji prikazuje mrežu klastera crvene boje jer su klasteri povezani isključivo negativnim vezama, tako je na ovom primeru i biće na svakom u daljem delu izveštaja.

clustered2.txt

Klasterabilna	Ukupno klastera	Klasteri koalicije	Klasteri koji nisu koalicije	Broj čvorova	Broj pozitivnih linkova	Broj negativnih linkova
Da	6	6	0	18	14	5



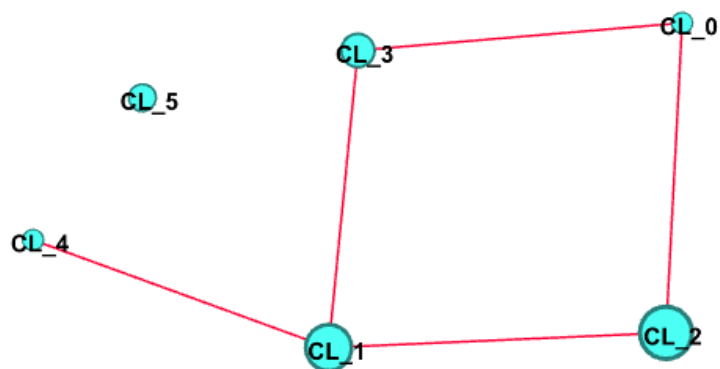
Slika 3

Na slici 3 prikazano je kako izgleda graf koji se analizira. Kao i u prethodnom slučaju veličina čvora je direktno proporcionalna stepenu čvora, a boja veze zavisi od znaka same veze.

Mreža je klasterabilna tako da ni ovde nema problema sa negativnim linkovima, a klasterisanjem mreže se dobija ukupno 6 klastera.

Na slici 4 prikazano je kako izgleda mreža klastera i u ovom primeru a kao i u prethodnom veličina čvora zavisi od broja čvorova unutar samog klastera. Gigantska komponenta je označena sa CL_2 i sadrži 6 čvorova u sebi, dok je ona na slici 1 uokvirena plavom bojom.

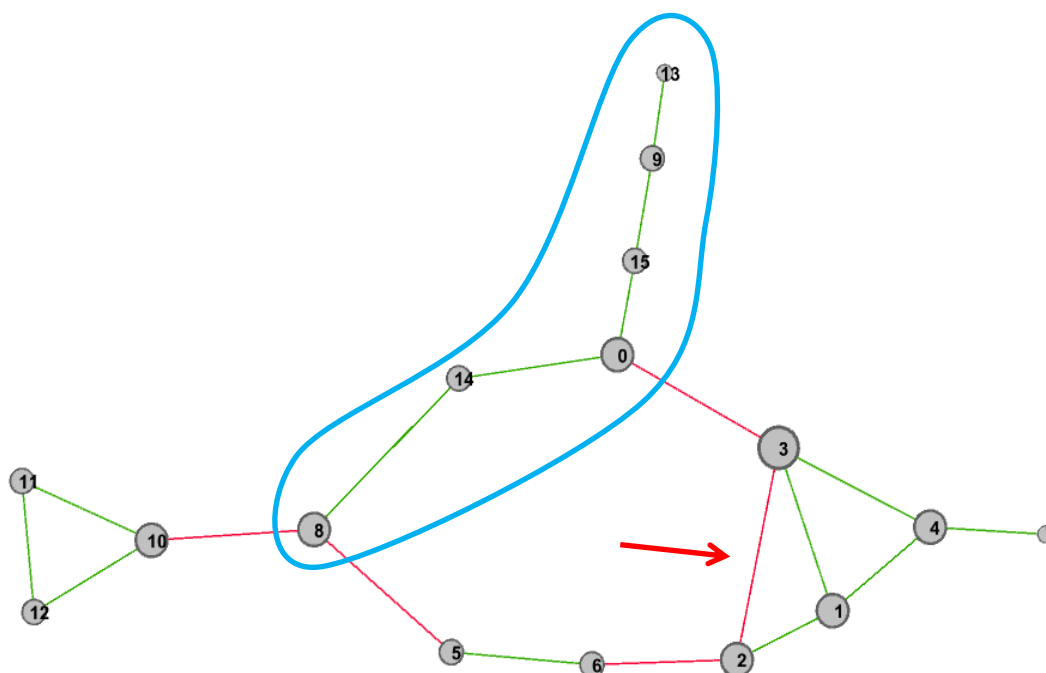
Takođe u ovom primeru vidimo da implementirani algoritam radi i kada graf nije povezan tj. na slici 4 vidimo da u mrežu postoji čvor koji sam čini jednu komponentu, čvor CL_5 koji predstavlja čvorove 11 i 16 u ulaznom grafu.



Slika 4

nonclustered1.txt

Klasterabilna	Ukupno klastera	Klasteri koalicije	Klasteri koji nisu koalicije	Broj čvorova	Broj pozitivnih linkova	Broj negativnih linkova
Ne	4	3	1	16	14	5

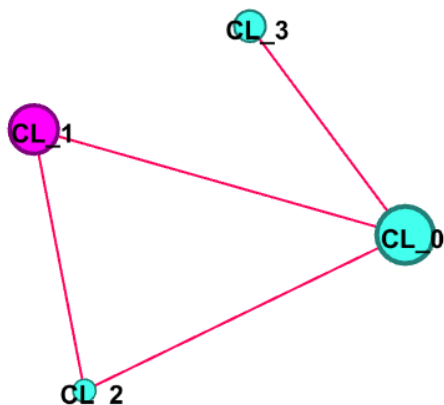


Slika 5

Na slici 5 je prikazano kako izgleda analizirani graf, ideja vizuelizacije je ista kao i u prethodna dva primera.

Mreža nije klasterabilna. Klasterisanjem ulaznog grafa dolazimo do ukupno 4 klastera od kojih 1 nije koalicija. Da bi mreža postala klasterabilna potrebno je izbrisati samo jednu vezu i to negativnu vezu između čvorova 2 i 3. Pomenuta veza je na slici obeležena strelicom.

Na slici 6 prikazana je mreža klastera. Kao i u prethodna dva slučaja veličina čvora zavisi od broja čvorova unutar klastera, međutim ovde boja čvora zavisi od toga da li je klaster i koalicija. Klasteri koji su koalicija prikazani su plavom bojom dok je klaster koji nije koalicija tj. u sebi sadrži negativan link prikazan nijansom ljubičaste boje.

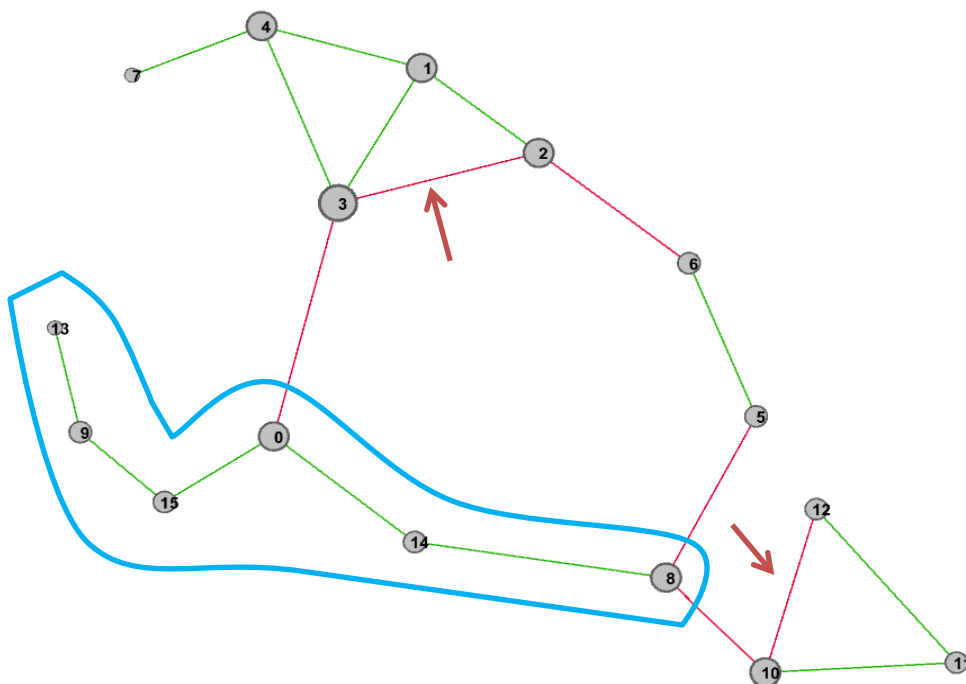


Slika 6

Gigantska komponenta u grafu je klaster CL_0, koji u sebi sadrži 6 čvorova, tj. taj klaster je na slici 5 uokviren plavom bojom.

nonclustered2.txt

Klasterabilna	Ukupno klastera	Klasteri koalicije	Klasteri koji nisu koalicije	Broj čvorova	Broj pozitivnih linkova	Broj negativnih linkova
Ne	4	2	2	16	13	6

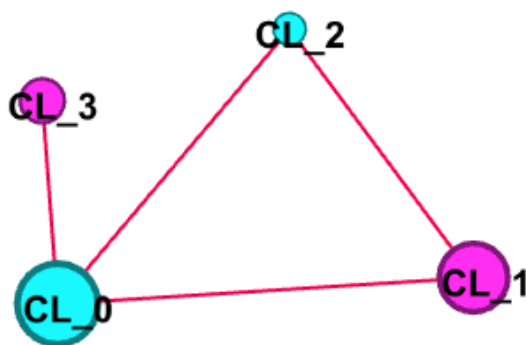


Slika 7

Na slici 7 prikazano je kako izgleda graf koji se analizira, pravila vizuelizacije su ista kao i u prethodnim primerima.

Mreža nije klasterabilna. Klasterisanjem ulaznog grafa dolazimo do ukupno 4 klastera od kojih 2 nisu koalicije. Da bi mreža postala klasterabilna treba izbrisati dve veze i to negativne veze između čvorova 2 i 3 i između čvorova 10 i 12, po jednu iz svakog klastera koji nije koalicija. Veze koje kvare klasterabilnost su na slici 7 obeležene strelicama.

Na slici 8 prikazana je mreža klastera. Kao i u prethodnom slučaju boja čvora zavisi od toga da li je klaster koalicija.

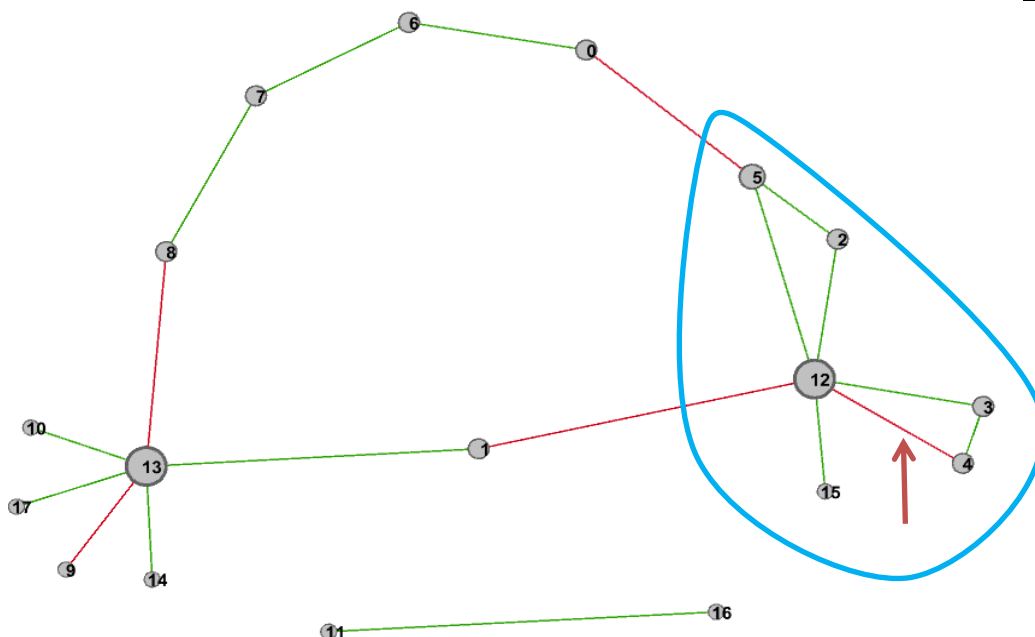


Slika 8

Gigantska komponenta u grafu je klaster CL_0 koji u sebi sadrži 6 čvorova i taj klaster je obeležen na slici 7 plavom bojom.

nonclustered3.txt

Klasterabilna	Ukupno klastera	Klasteri koalicije	Klasteri koji nisu koalicije	Broj čvorova	Broj pozitivnih linkova	Broj negativnih linkova
Ne	5	4	1	18	14	5

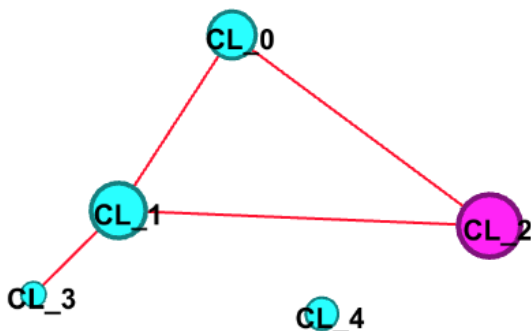


Slika 9

Na slici 9 je prikazano kako izgleda graf koji se analizira. Graf je prikazan na prethodno već poznat način, gde boja veza zavisi od znaka veze a veličina čvora od njegovog stepena.

Mreža nije klasterabilna. Klasterisanjem ulaznog grafa dolazi se do ukupno 5 klastera od kojih samo 1 nije koalicija. Da bi mreža postala klasterabilna treba izbrisati samo jednu vezu i to negativnu vezu između čvorova 4 i 12. Pomenuta veza je naznačena na slici 9.

Na slici 10 prikazana je mreža komponenti analiziranog grafa. Vizuelizacija se radi po već utvrđenim pravilima.



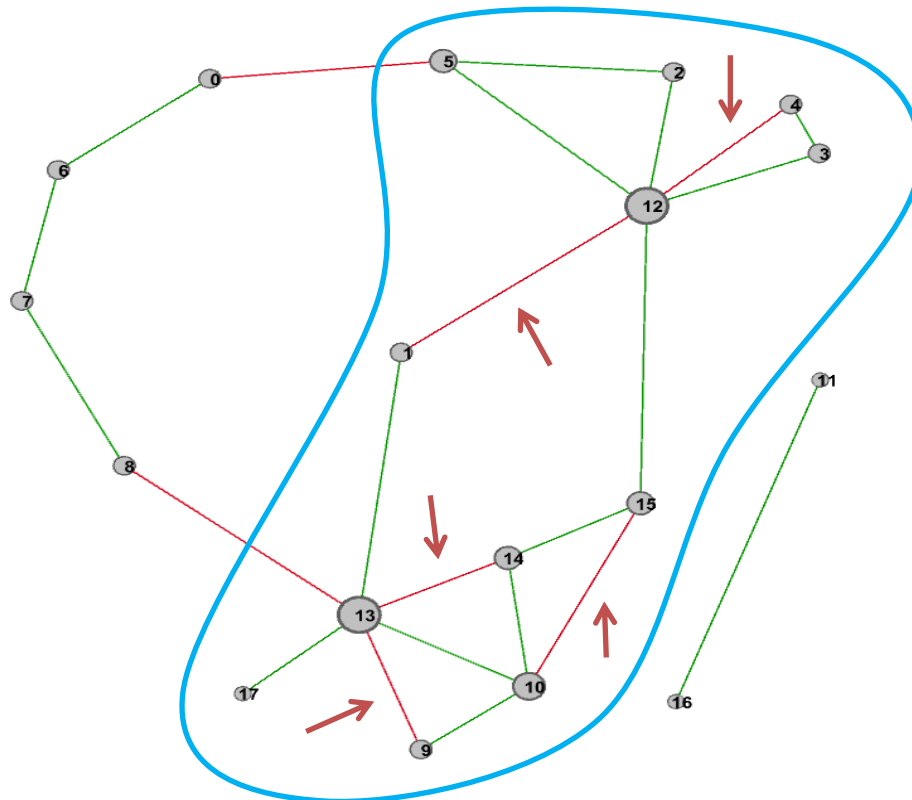
Slika 10

Gigantska komponenta je klaster označen sa CL_2 na slici 10, takođe je i uokviren na slici 9. Gigantska komponenta u sebi sadrži 6 čvorova, a u ovom primeru je to upravo i jedina komponenta koja nije koalicija.

nonclustered4.txt

Klasterabilna	Ukupno klastera	Klasteri koalicije	Klasteri koji nisu koalicije	Broj čvorova	Broj pozitivnih linkova	Broj negativnih linkova
Ne	3	2	1	18	16	7

Na slici 10 vidimo kako izgleda graf koji se u ovom slučaju analizira. Način vizuelizacije je isti kao i prethodnim primerima.



Slika 11

Mreža nije klasterabilna. Klasterisanjem ulaznog grafa dolazi do ukupno 3 klastera od kojih su 2 koalicije a 1 nije. Da bi mreža postala klasterabilna treba izbrisati 5 veza i to negativne veze između čvorova 1 i 12, 4 i 12, 9 i 13, 10 i 15, 13 i 14. Sve veze koje treba izbrisati su označene na slici 11

Na slici 12 prikazana je mreža komponenti analiziranog grafa. Vizuelizacija se radi po već utvrđenim pravilima.



Slika 12

Gigantska komponenta je klaster označen sa CL_1 na slici 12, takođe je i uokviren na slici 9. Gigantska komponenta u sebi sadrži čak 12 čvorova i ona je jedini klaster koji nije koalicija.

Random mreže

U nastavku će biti prikazane i analizirane neke slučajno generisane mreže koje su rađene na predavanjima u okviru kursa.

Mreže koje će se koristiti su generisane pomoću već pomenutih klasa koje se nalaze u paketu `networks` i svim mrežama možemo ručno podešavati broj čvorova, broj veza ili verovatnoću da su neka dva čvora povezana u zavisnosti od modela koji je koristi, takođe kod svakog modela možemo zadavati verovatnoću za kreiranje negativne veze između čvorova.

Takođe za ove mreže su primenjene i metrike o kojima je bilo reči u uvodnom delu.

Analiza počine od Erdos-Renyi modela i ide ka složenijim mrežama.

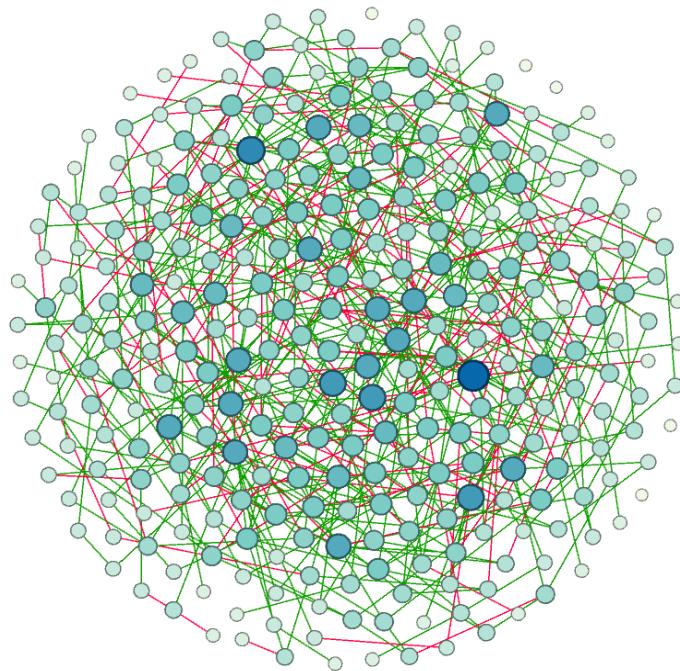
Erdos-Renyi model

Erdos-Renyi model slučajnih grafova nije model koji je specijalno predviđen za rad sa označenim grafovima tako da u ovom zadatku klasa koja poziva generator mreže u `Supplier` objekat predviđen za kreiranje grana u grafu zadaje verovatnoću za kreiranje negativne grane unutar grafa.

```
Supplier<MarkedEdge> edgeFactory = new Supplier<MarkedEdge>() {  
    @Override  
    public MarkedEdge get() {  
        Random rnd = new Random();  
        return new MarkedEdge(rnd.nextDouble() <  
            negativeLinkProbability ? Mark.NEGATIVE : Mark.POSITIVE);  
    }  
};
```

Parametar *negativeLinkProbability* je prosleđen kroz konstruktor.

Mreža koja će biti analizirana u nastavku je konstruisana sa sledećim parametrima: broj čvorova je 300, broj veza je 600 i verovatnoća kreiranja negativne veze je 25%

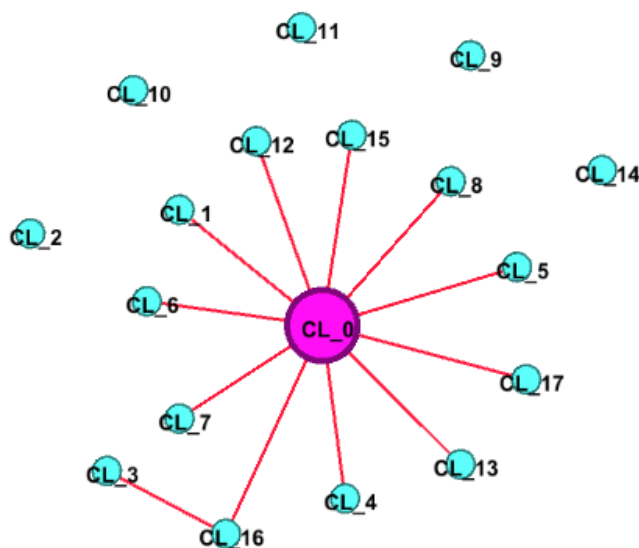


Slika 13

Na slici 13 prikazano je kako izgleda izgenerisana random Erdos-Renyi mreža. Veze sa pozitivnim znakom prikazane su zelenom bojom, a veze sa negativnim znakom prikazane su crvenom bojom kao i do sada, veličina čvora zavisi od stepena čvora ali sada i boja čvora zavisi od stepena čvora, sa povećanjem stepena plava boja čvorova postaje tamnija. U ovom slučaju kao i kod narednih modela čvorovi su vizuelizovani ovako radi preglednosti.

Klasterabilna	Ukupno klastera	Klasteri koalicije	Klasteri koji nisu koalicije	Broj čvorova	Broj pozitivnih linkova	Broj negativnih linkova
Ne	18	17	1	300	441	161

Mreža nije klasterabilna. Klasterisanjem mreže dobija se 18 klastera od kojih samo jedan nije koalicija i to je najveći klaster, koji predstavlja i gigantsku komponentu ovog grafa. Da bi mreža postala klasterabilna potrebno je izbrisati čak 142 negativne veze od ukupno 161 koliko ih se nalazi u celom grafu. Od 17 koalicija u okviru generisanog grafa čak 15 njih je veličine jednog čvora. Na slici 14 je prikazano kako izgleda mreža komponenti ovog grafa.



Slika 14

Na slici su koalicije obojene nijansom plave boje, a klasteri koji nisu koalicije su obojeni nijansom ljubičaste. Vidimo da je jedini klaster koji nije koalicija upravo gigantska komponenta grafa. Gigantska komponenta se sastoji od 281 čvora, što je preko 90% čvorova u grafu.

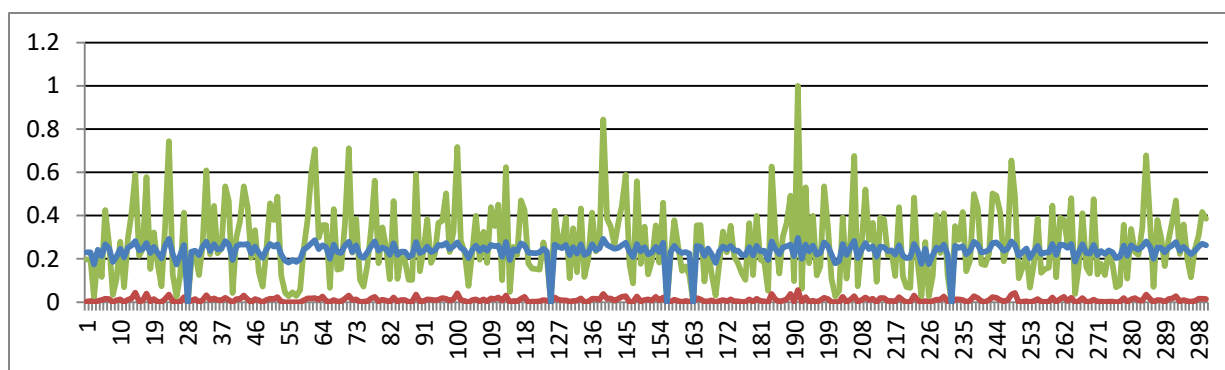
Dijametar u gigantskoj komponenti je 8, dok je prosečan koeficijent klasterisanja oko 0.02 što je i očekivano jer Erdos-Renyi mreže ne karakteriše visok stepen klasterisanja. Prosečan stepen klasterisanja celog grafa je još niži i iznosi oko 0.018.

Čvor sa najvećim stepenom klasterisanja je čvor 44 koji ima samo dve veze i čini trougao sa druga dva grafa, obe njegove veze su pozitivne. Njegov koeficijent klasterisanja je 1.

Small-world koeficijent nema svrhe meriti na nivou celog grafa pošto graf nije povezan, a na nivou gigantske komponente on iznosi približno 2.064 a efikasnost grafa je približno 0.13. Pošto je $\log(n) \approx 2.477$ može se reći da gigantska komponenta poseduje small-world osobinu.

Čvor 190 ima najveće vrednost i za *betweenness* i za *closeness* centralnost, dok se ostali čvorovi sa najvećim vrednostima ovih centralnosti manje više poklapaju.

Metrike centralnosti prikazane su na dijagramu ispod. Na dijagramu crvenom bojom je predstavljena *betweenness* centralnost, plavom bojom je predstavljena *closeness* centralnost, a zelenom bojom je predstavljena *eigenvector* centralnost.



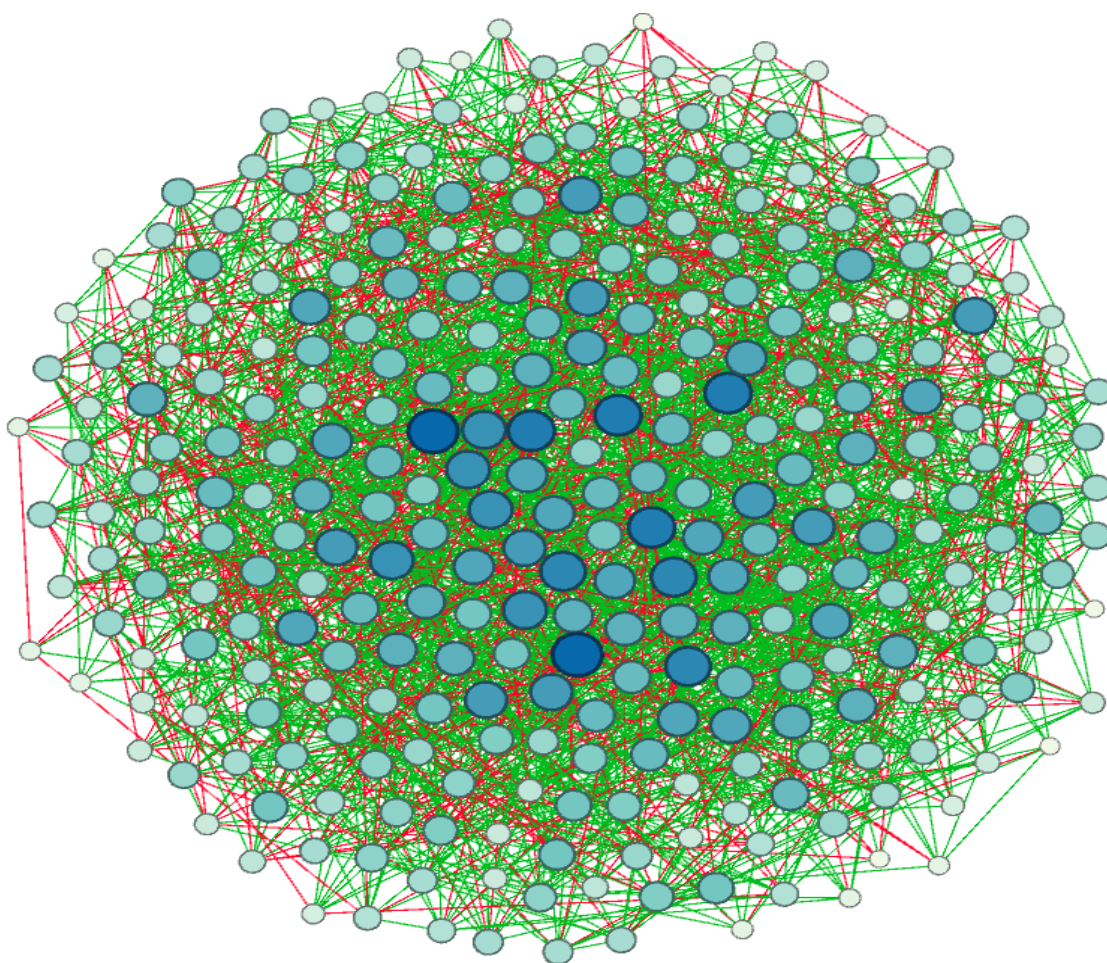
Grafikon je eksportovan iz alata Gephi, vrednosti koje daje alat se u potpunosti poklapaju sa vrednostima koje daje i sama implementacija u okviru ovog projekt.

Gilbert model

Kao ni Erdos-Renyi model, Gilbert model nije inicijalno dizajniran kao označen graf, međutim i ovde za potrebe zadatka kao i kod prethodnog modela u konstrukciju prosleđujemo verovatnoću kreiranja negativne veze koja se kroz `Supplier` prenosi u generator.

Mreža koja će biti analizirana generisana je na osnovu sledećih parametara: broj čvorova je 300 kao i kod prethodnog modela, međutim ovde se ne zadaje broj veza koje treba ostvariti nego verovatnoća da se bilo koja veza ostvari, verovatnoću zadajemo 5%, a verovatnoća za kreiranje negativnih linkova biće 30%.

Parametri su uzeti kao najidealniji za prikaz mreže tako da se na njoj može nešto zapravo i videti a ne da to samo bude milion linija isprepletanih međusobno od kojih se neće videti ništa.



Slika 15

Na slici 15 je prikazan izgled navedenog grafa, vizuelizacija je odrađena po istim parametrima kao i kod prethodnog modela.

Klasterabilna	Ukupno klastera	Klasteri koalicije	Klasteri koji nisu koalicije	Broj čvorova	Broj pozitivnih linkova	Broj negativnih linkova
Ne	1	0	1	300	1596	673

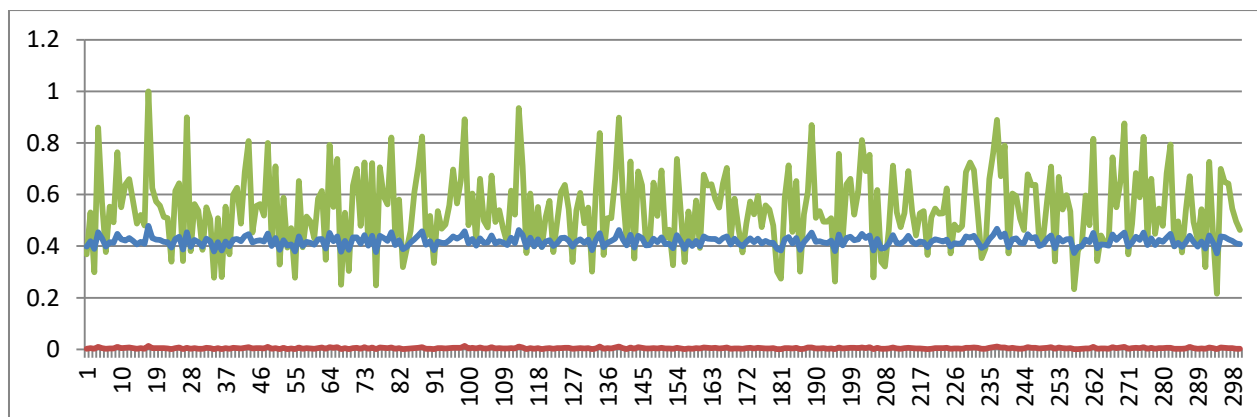
Mreža nije klasterabilna. Klasterisanjem ulazne mreže dobijamo samo jedan klaster, ako bi se procenat negativnih linkova povećao tako bi se povećao i broj klastera ali ne značajno. Jedini klaster koji je dobijen nije koalicija i on predstavlja i gigantsku mrežu grafa kao i ceo graf uostalom.

Pošto se na osnovu analite grafa vidi da mrežu komponenti čini samo jedan čvor onda nema potrebe prikazivati mrežu.

Prosečan stepen klasterisanja je približno 0.048, dok je čvor sa najvećim stepenom klasterisanja čvor 33 sa stepenom približno 0.14. Ovakve mere su očekivane pošto kao ni Erdos-Renyi model ni Gilbertov model ne generiše grafove sa visokim koeficijentom klasterisanja.

Dijametar grafa je 4. Graf je dobro povezan. Small-world koeficijent na nivou gigantske komponente kao i celeog grafa iznosi približno 1.2 a efikasnost grafa je približno 0.226. Pošto je $\log(n) \approx 2.477$, n je broj čvorova, može se reći da graf poseduje small-world osobinu, small-world koeficijent je duplo manji.

Kada se posmatraju metike centralnosti izvodi se zaključak kao i kod Erdos-Renyi modela, sve tri metrike centralnosti favorizuju iste čvorove sa malim odstupanjima. Uglavnom se kod sve tri metrike isti čvorovi nalaze na vrhovima liste samo su im mesta u zavisnosti od centralnosti izmešana. Closeness centralnost favorizuje čvor 16, kao i eigenvector centralnost, dok betweenness centralnost čvor 16 stavlja na drugo mesto a favorizuje čvor 98

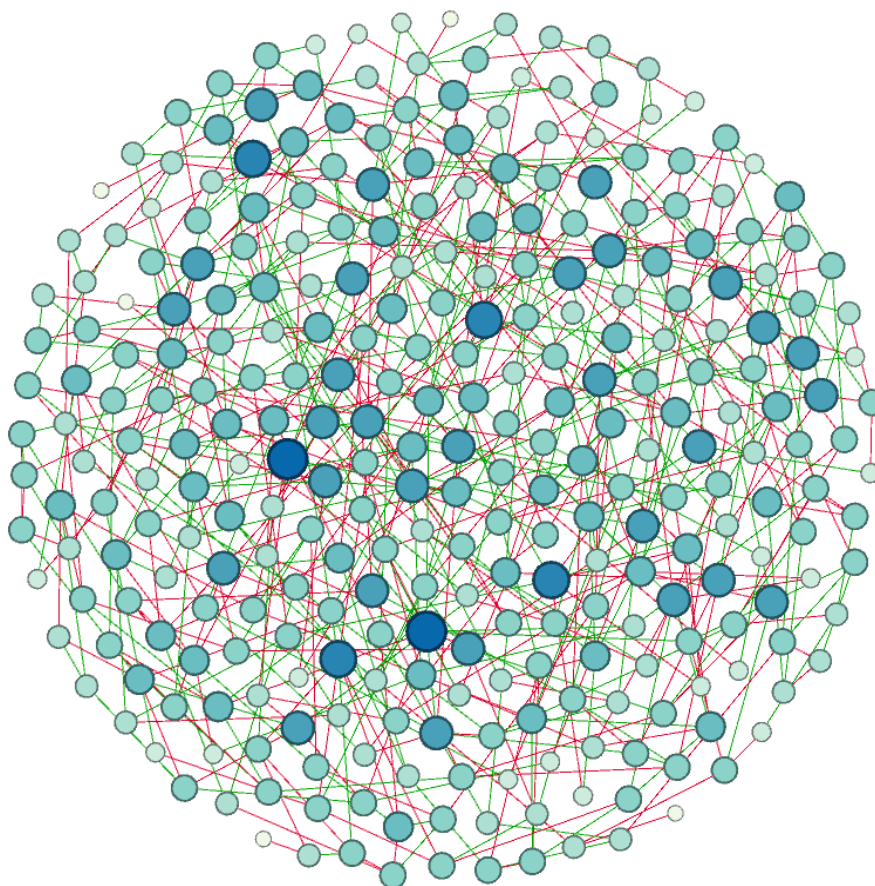


Na grafikonu je zelenom bojom prikazana eigenvector, plavom closeness a crvenom betweenness centralnos, kao i kod Erdos-Renyi modela csv fajl je generisan direktno iz alata Gephi jer se vrednosti dobijene u alatu podudaraju sa vrednostima koje generiše i sam program.

Watts-Strogatz Model

Kao ni Erdos-Renyi i Gilbert modeli ni Watts-Strogatz model nije inicijalno dizajniran kao označen graf, međutim i ovde za potrebe zadatka kao i kod prethodnih modela u konstrukciju prosleđujemo verovatnoću kreiranja negativne veze koja se kroz Supplier prenosi u generator.

Kod ovog modela imamo nešto drugačije parametre od onih kod dosadašnjih modela. Dakle parametri po kojima je dizajnirana mreža koja se analizira su sledeći: broj čvorova će biti 300 kao i do sada, međutim ovde prosleđujemo parametar k koji označava inicijalan broj veza svakog čvora u inicijalno kreiranom grafu, k će biti 4, parametar p koji označava verovatnoću promene veze biće 20% odnosno 0.2 a verovatnoća kreiranja negativne veze ovde će biti 50% odnosno 0.5. Parametri su izabrani tako da se kreira koliko toliko pregledna mreža i da na njoj mogu da se demonstriraju neke osobine ovog modela.

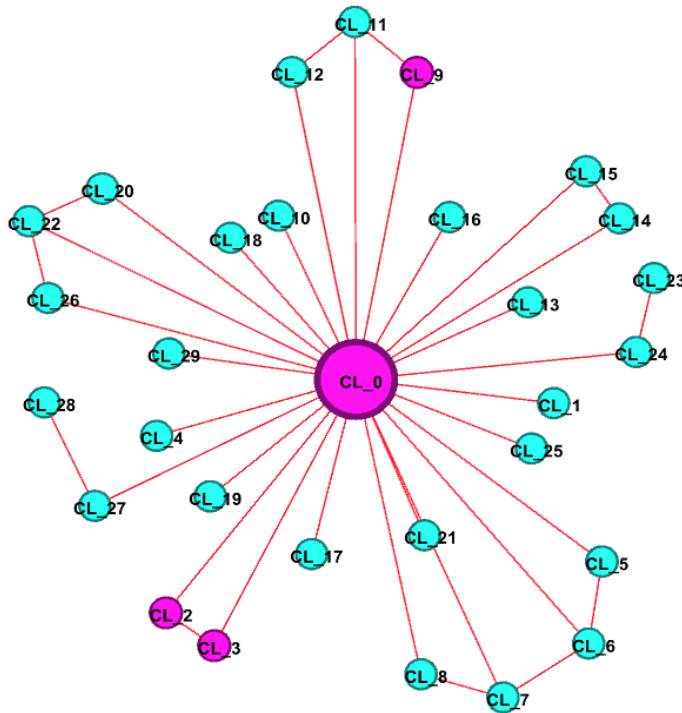


Slika 16

Na slici 16 prikazano je kako izgleda izgenerisani graf. Parametri vizuelizacije su isti kao i do sada i važe ista pravila.

Klasterabilna	Ukupno klastera	Klasteri koalicije	Klasteri koji nisu koalicije	Broj čvorova	Broj pozitivnih linkova	Broj negativnih linkova
Ne	30	26	4	300	303	297

Mreža nije klasterabilna. Klasterisanjem mreže dolazimo do ukupno 30 klastera od kojih su njih 26 koalicije a 4 nisu. Od 26 koalicija čak 18 se sastoji iz samo jednog čvora. Da bi mreža postala klasterabilna potrebno je ukloniti 198 negativnih linkova od ukupno 297 koliko je izgenerisano. Pokretanjem programa moguće je videti i koje linkove tačno treba izbrisati.



Slika 17

Na slici 17 prikazano je kako izgleda mreža komponenti generisanog grafa. Pravila vizuelizacije ista su kao kod Erdos-Renyi modela.

Vidimo da gigantska komponenta grafa nije koalicija i ona se sastoji iz ukupno 249 čvorova. Gigantska komponenta ima dijametar 9, prosečan koeficijent klasterisanja 0.143 a čvor sa najvećim stepenom klasterisanja je čvor 19 i njegov stepen iznosi 1.

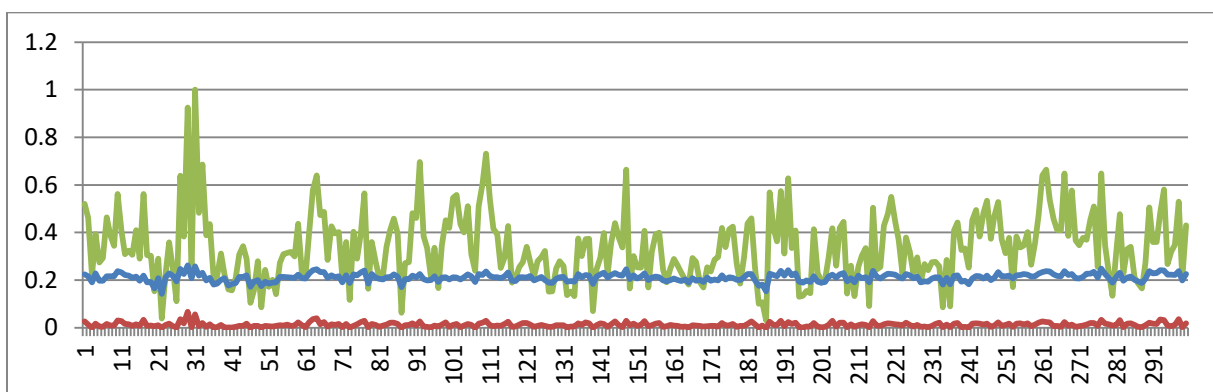
Small-world koeficijent gigantske komponente ima vrednost oko 2.399, što je gotovo identično vrednosti $\log(n)$ koja iznosi približno 2.396 što dovodi do zaključka da je small-world osobina izražena.

Efikasnost u gigantskoj komponenti 0.118.

Na nivou celog grafa prosečan stepen klasterisanja iznosi oko 0.134 što je nekoliko puta veće nego kod prethodno analiziranih modela, što je očekivano kod ovog modela budući da generiše grafovi se nešto većim koeficijentom klasterisanja. Na nivou celog grafa najveći stepen klasterisanja imaju čvorovi 2 i 106 a to je 1.

Small-world koeficijent na nivou celog grafa ima vrednost 2.38 što je značajno blizu vrednosti $\log(n)$ koja iznosi nešto manje od 2.48. Zaključujemo da graf poseduje small-world osobinu i to baš onako kako je očekivano od ovakvog tipa grafa. Efikasnost celog grafa je 0.118.

Kada se gledaju metrike centralnosti zaključuje se da metrike favorizuju manje više iste čvorove, sve metrike favorizuju čvor 28 kao najbitniji čvor u grafu.

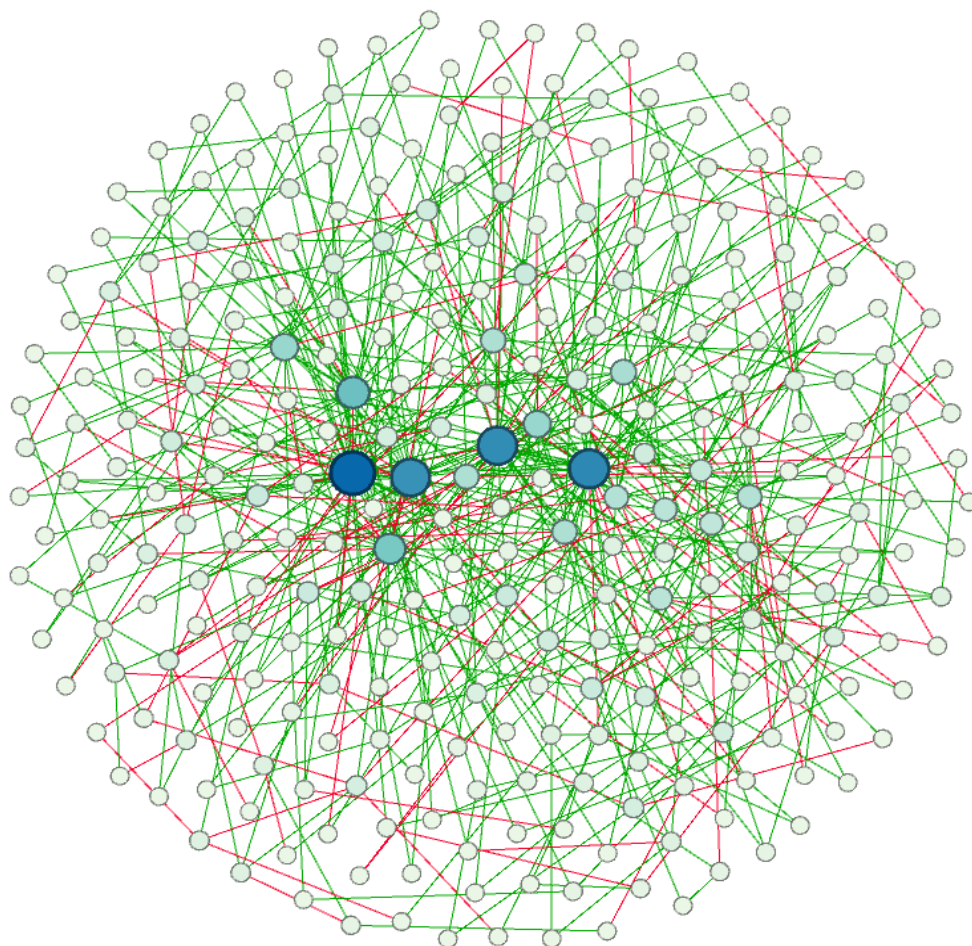


Na grafikonu su kao i do sada zelenom bojom prikazane eigenvector centralnosti, plavom bojom closeness centralnosti a crveom bojom betweenness centralnosti. Posmatrajući grafikon može se videti da metrike prate jedna drugu sa povremenim ali ne značajnim odstupanjem u vrhu svih metrika nalaze se isti čvorovi.

Barabasi-Albert model

Kao i kod prethodno analiziranih modela i kod Barabasi-Albert modela nailazimo na isti problem, model nije dizajniran za rad sa označenim grafovima tako da i ovde kao i u prethodnim slučajevima prilikom inicijalizacije zadajemo dodatni parametar koji predstavlja verovatnoću kreiranja negativne veze.

Graf koji će biti analiziran ima sledeće parametre: broj čvorova će biti 300 kao i do sada, međutim kod ovog modela potrebno je zadati i parametre inicijalne mreže koja se kasnije proširuje, broj čvorova inicijalne mreže biće 10, broj grana inicijalne mreže biće 20, a verovatnoća kreiranja negativne veze u inicijalnom grafu će biti 25%, kreira se Erdos-Renyi model, broj veza koje će se dodavati u čvoru prilikom proširenja biće 2 a verovatnoća da prilikom proširenja grafa bude kreirana negativna veza biće 20%.

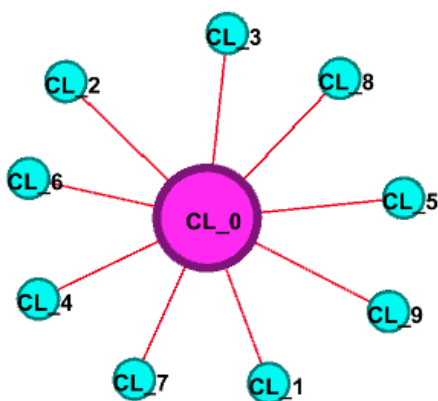


Slika 18

Na slici 18 dat je prikaz grafa koji se analizira i koji je generisan po navedenim parametrima, pravila vizuelizacije su ista kao do sad

Klasterabilna	Ukupno klastera	Klasteri koalicije	Klasteri koji nisu koalicije	Broj čvorova	Broj pozitivnih linkova	Broj negativnih linkova
Ne	10	9	1	300	472	124

Mreža nije klasterabilna. Klasterisanjem ulazne mreže dolazi se do 10 klastera od kojih su 9 koalicije a 1 nije. Od 9 koalicija čak 8 njih se sastoji od samo jednog čvora. Da bi mreža postala klasterabilna potrebno je izbrisati 106 negativnih veza od ukupno 124 koliko ih ima u celoj mreži.



Slika 19

Na slici 19 prikazano je kako izgleda mreža komponenti generisanog grafa. Pravila vizuelizacije su ista kao i kod prethodnih modela.

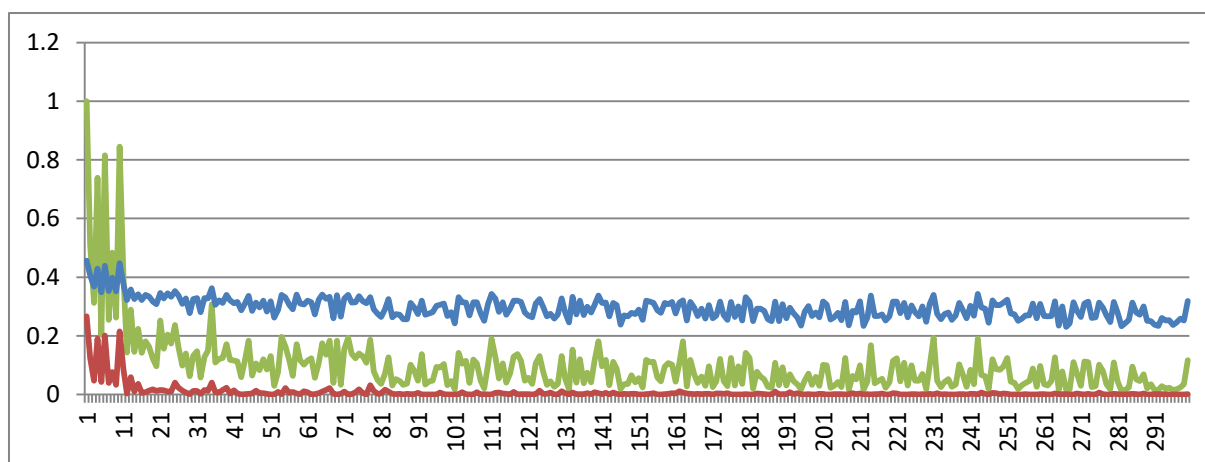
Vidimo da jedina komponenta koja nije koalicija je i najveća tj. gigantska komponenta mreže. Sastoji se od ukupno 290 čvorova, dakle samo 10 čvorova mreže se nalazi van gigantske komponente. Dijametar gigantske komponente je 6, prosečan koeficijent klasterisanja iznosi približno 0.08, a čvor sa najvećim stepenom klasterisanja je čvor 43 i on ima stepen 1 i on u gigantskoj komponenti gradi trougao gde su stranice trougla tri pozitivne veze. Small-world koeficijent gigantske komponente je približno 1.71, a

pošto $\log(n)$ iznosi približno 2.46 možemo reći da gigantska komponenta poseduje small-world osobinu. Efikasnost komponente je približno 0.16.

Na nivou celog grafa prosečan stepen klasterisanja iznosi približno 0.077, dok je čvor sa najvećim stepenom klasterisanja čvor 50, ima vrednost 1, on gradi trougao samo sa pozitivnim vezama. Stepem klasterisanja je duplo manji nego kod npr. Watts-Strogatz modela što je i očekivano.

Small-world koeficijent na nivou celog grafa iznosi približno 1.72 što je manje od vrednosti $\log(n)$ koja iznosi oko 2.48. Može se reći da mreža poseduje small-world osobinu. Efikasnost mreže je približno 0.16, dok je dijametar celog grafa 6. Vidimo da se ove mere na nivou celog grafa mahom poklapaju sa merama na nivou gigantske komponente što je i normalno za ovaj model jer u grafu postoji nekoliko čvorova koji imaju velike stepene i povezani su sa većinom ostalih čvorova. To su uglavnom čvorovi koji su se nalazili u inicijalnoj Erdos-Renyi mreži što će se još jasnije videti na metrikama centralnosti.

Kada govorimo o metrikama centralnosti na dijagramu ispod vidimo da metrike favorizuju iste čvorove, zapravo metrike u pet najbitnijih čvorova svrstavaju istih pet čvorova. Po svim metrikama najbitniji čvor je čvor sa ID brojem 0.



Kao što je već rečeno na ovom dijagramu se jasnije vidi a to je da čvorovi sa manjim id-ovima imaju i veće vrednosti ovih metrika a to su upravo čvorovi inicijalne mreže i oni predstavljaju kritične tačke ove mreže, njihovim uklanjanjem mreža bi se raspala, a to je i zapravo odlika Barabasi-Albert modela.

Kao i do sada na dijagramu je closeness centralnost prikazana plavom linijom, betweenness crvenom a eigenvector centralnost zelenom linijom.

Realne mreže

U nastavku će biti analizirane realne mreže Slashdot, Epinions i WikiRfA redom. Mreže su preuzete sa [Stanford Large Network Dataset Collection](#).

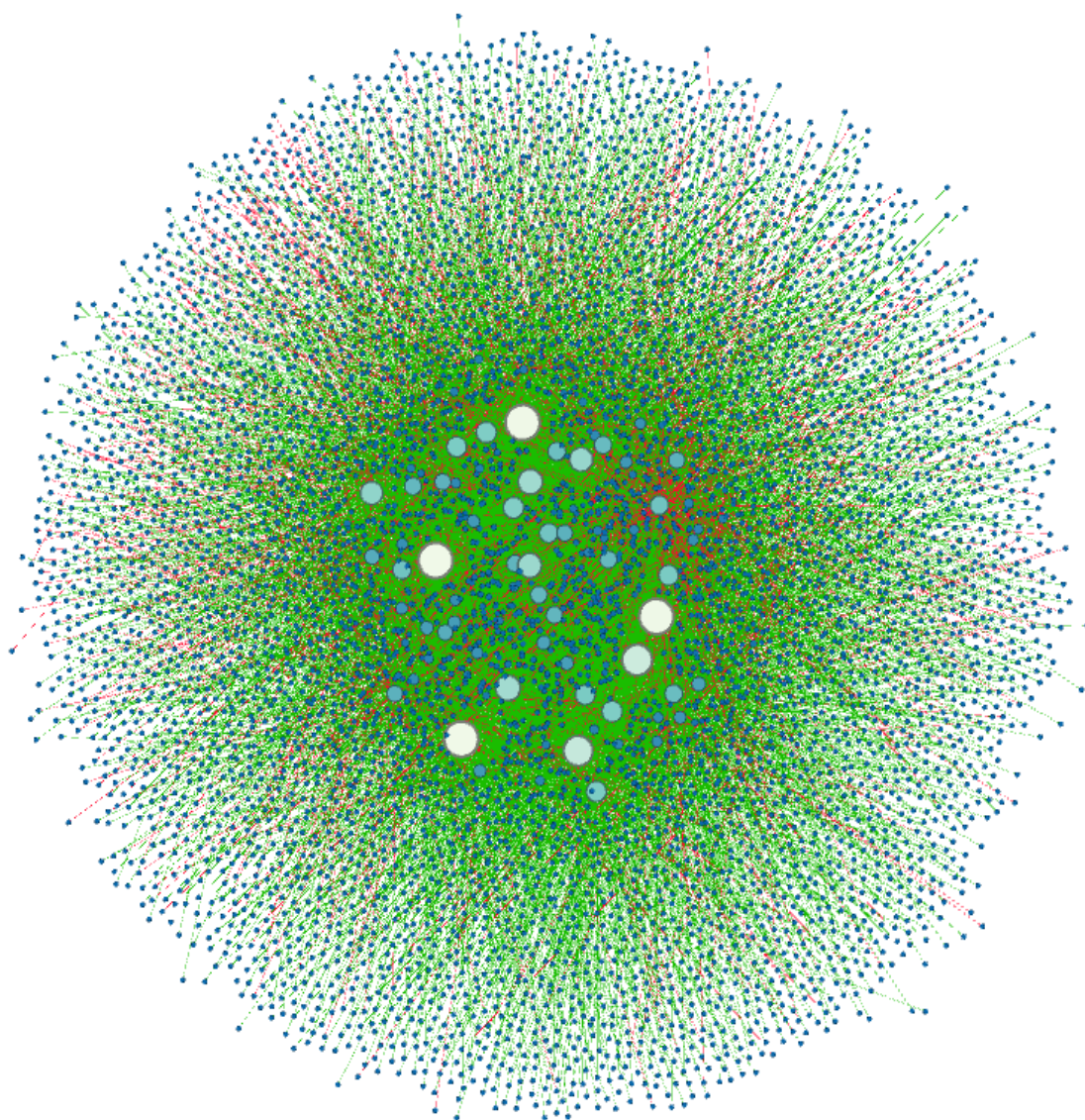
Mreže će najpre biti učitane u program iz odgovarajućih txt fajlova, zatim će biti analizirane i eksportovane u GraphML format radi vizuelizacije u alatu Gephi. Zbog mogućnosti alata a i samog računara na kom se radi analiza plus zbog trajanja učitavanja mreže neće biti učitavane cele mreže već samo delovi tj. 12 000 linija svakog fajla.

Sve mreže koje će se analizirati su i usmerene mreže, međutim usmerenja veza ignorišemo tako što ako su dva korisnika povezana sa dve pozitivne veze veza u mreži će ostati pozitivna dok će u bilo kojoj drugoj kombinaciji veza biti negativna, dakle dajemo prioritet negativnoj vezi.

Slashdot mreža

Slashdot je veb sajt sa vestima iz oblasti tehnologije koja je poznata po svojoj specifičnoj korisničkoj zajednici. Veb stranica sadrži pre svega vesti orijentisane ka tehnologiji i te vesti postavljaju korisnici dok ih administratori odobravaju i ocenjuju. Slashdot je 2002. godine uveo funkcionalnost pod imenom *Slashdot Zoo* koja omogućava svojim korisnicima da druge korisnike označe kao svoje prijatelje ili „neprijatelje“. U nastavku se analizira Slashdot mreža objavljena u novembru 2008 godine i mreža sadrži prijatelj/neprijatelj veze između korisnika ovog sajta u trenutku objavljivanja mreže. Mreža je usmerena ali u projektu se usmerenje ignorišu.

Originalna mreža ima 77350 čvorova i 516575 veza, ali se ovde ograničavamo na prvih 12 000 linija fajla kao što je prethodno rečeno.

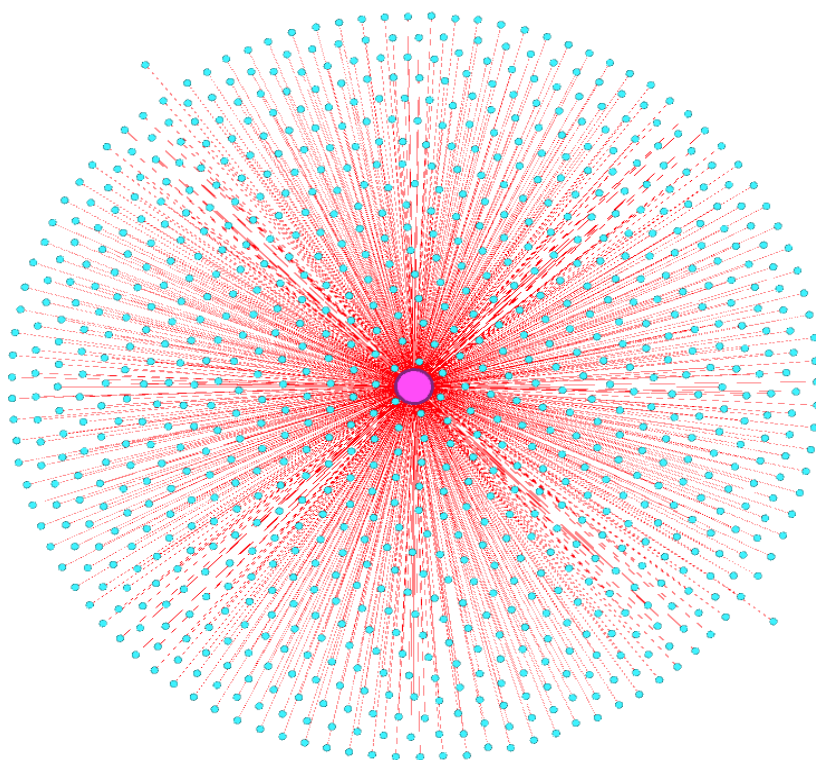


Slika 20

Na slici 20 prikazano je kako izgleda deo mreže koji se analizira. Vizuelizacija se odvija po istim parametrima kao i kod vizuelizacije random mreža, jedina razlika je što su sada čvorovi sa većim stepenom predstavljeni svetlijom nijansom boje a oni sa manjim su tamnom nijansom plave, razlog invertovanju boja je bolja vidljivost, a takođe je interval za veličinu čvorova smanjen iz istog razloga.

Klasterabilna	Ukupno klastera	Klasteri koalicije	Klasteri koji nisu koalicije	Broj čvorova	Broj pozitivnih linkova	Broj negativnih linkova
Ne	906	905	1	6281	9959	1665

Mreža nije klasterabilna. Klasterisanjem ulazne mreže dolazi se do 906 klastera od kojih su 905 koalicije a samo 1 nije. Od 905 koalicija čak 904 od njih se sastoji iz samo jednog čvora. Da bi analizirani deo mreže postao klasterabilan potrebno je izbrisati 657 negativnih veza od ukupno 1665 koliko ukupno deo mreže ima.



Slika 21

Na slici 21 prikazano je kako izgleda mreža komponenti dela mreže koji se analizira. Pravila vizuelizacije su ista kao i do sada.

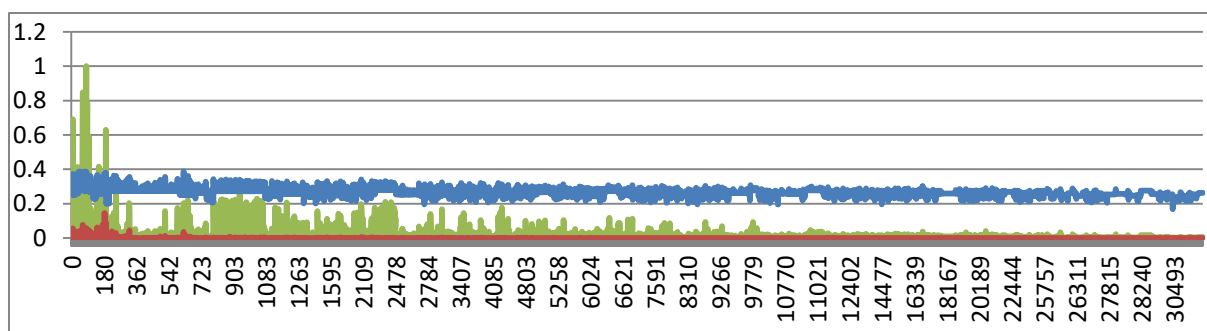
Vidimo da jedina komponenta koja nije koalicija jeste i gigantska komponenta grafa. Ona se sastoji od 5373 čvora. Prosečan koeficijent klasterisanja u gigantskoj komponenti je približno 0.096, a postoji više čvorova koji imaju koeficijent klasterisanja 1 a to su čvorovi koji su temena trouglova koji za stranice imaju tri pozitivne veze

Zbog memorijskog prostora na java heap-u

small-world koeficijent nije računat ni za gigantsku komponentu kao ni za ceo graf. Dijametar u gigantskoj komponenti je 8.

Na nivou celog grafa prosečan stepen klasterisanja iznosi približno 0.294, a kao i kod gigantske komponente dosta čvorova ima stepen klasterisanja 1 i to su čvorovi koji svojim pozitivnim vezama grade trouglove sa drugim čvorovima. Dijametar na nivou mreže je 8 isto kao i kod gigantske komponente.

Na dijagramu ispod prikazane su metrike centralnosti za celu mrežu. Kao i do sada eigenvector centralnost prikazana je zelenom linijom, closeness centralnost prikazana je plavom a betweenness crvenom linijom.

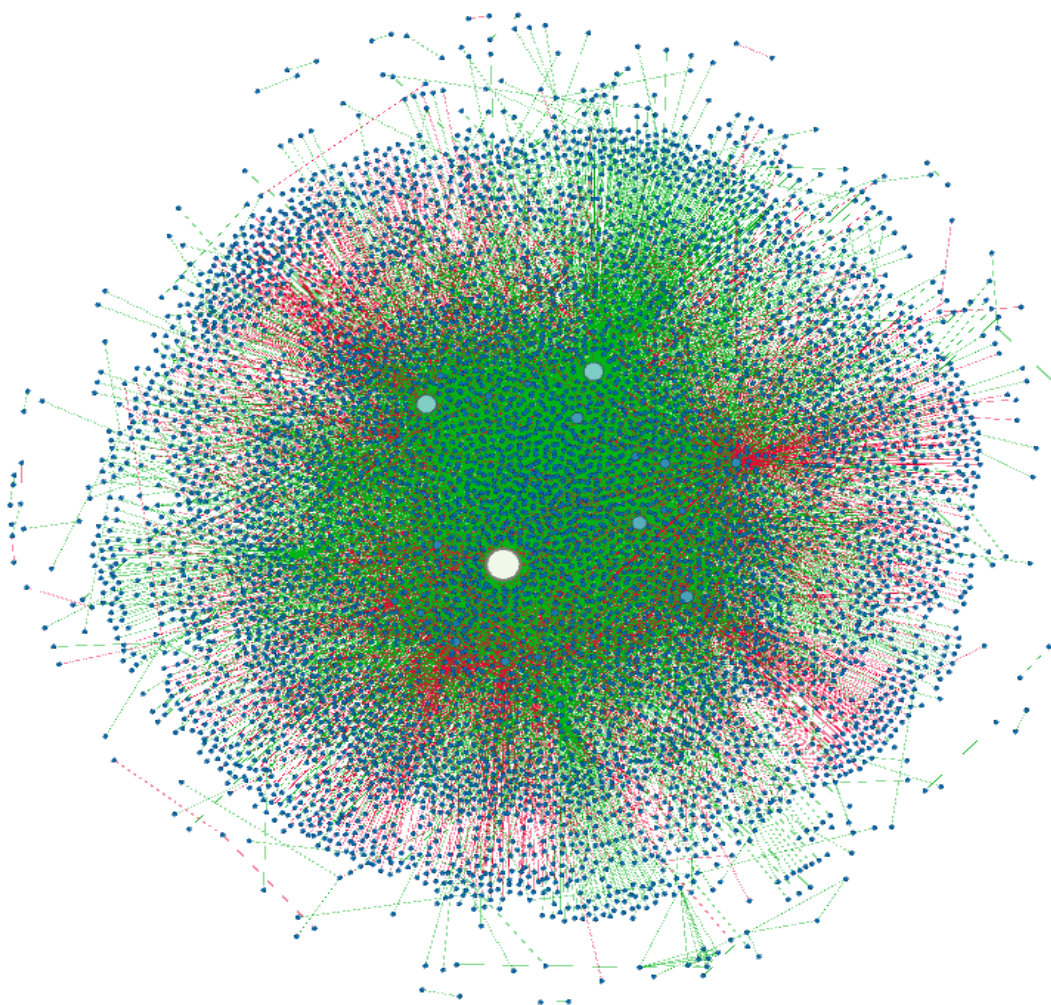


Vidimo da se closeness i eigenvector centralnosti slažu tj. da favorizuju iste čvorove, takođe nema nekog velikog odstupanja ni kod betweenness centralnosti. Najveće vrednosti za metrike centralnosti imaju upravo čvorovi koji imaju i najveći stepen u mreži, tj. čvorovi koji imaju najviše uspostavljenih odnosa među korisnicima, ovakav rezultat je očekivan jer su ti čvorovi najbitniji za propagiranje informacija kroz mrežu.

Epinions mreža

Epinions je potrošačka mreža na internetu gde su korisnici ove mreže mogli da označe da li veruju nekom drugom korisniku ili ne. Sve veze poverenja su bile u interakciji međusobno tako da je kreirana mreža poverenja. U ovom delu biće analiziran deo jedne takve mreže.

Originalna mreža se sastoji od 131828 čvorova i 841372 veza između njih, međutim kao što je već rečeno u ovom delu će iz poznatih razloga biti analizirano samo prvih 12 000 linija fajla. Takođe i ova mreža je usmerena kao i prethodna ali će biti posmatrana kao neusmerena mreža na isti način kao i prethodna mreža.



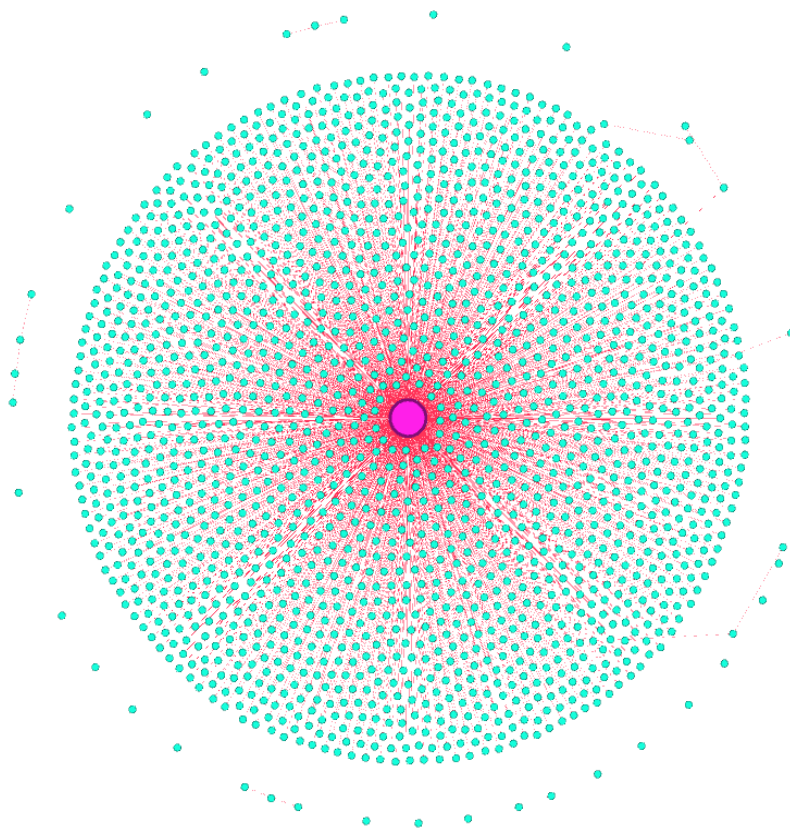
Slika 22

Na slici 22 prikazano je kako izgleda deo mreže koji se analizira, način vizuelizacije je isti kao i kod Slashdot mreže.

Klasterabilna	Ukupno klastera	Klasteri koalicije	Klasteri koji nisu koalicije	Broj čvorova	Broj pozitivnih linkova	Broj negativnih linkova
Ne	1874	1873	1	6536	9160	2737

Mreža nije klasterabilna. Klasterisanjem ulazne mreže dolazi se do 1874 klastera od kojih su 1873 koalicije a samo jedan nije. Od ukupno 1873 klastera koji su koalicije postoji čak 1842 koalicije sa samo jednim čvorom. Da bi mreža postala klasterabilna potrebno je izbrisati 605 negativnih veza od ukupno 2737 koliko ih se nalazi u analiziranoj podmreži.

Na slici 23 prikazano je kako izgleda mreža komponenti posmatranog grafa. Mreža



Slika 23

komponenti se sastoji iz 1874 čvora. Vidimo da je gigantska komponenta i jedini klaster koji nije koalicija u analiziranom delu mreže.

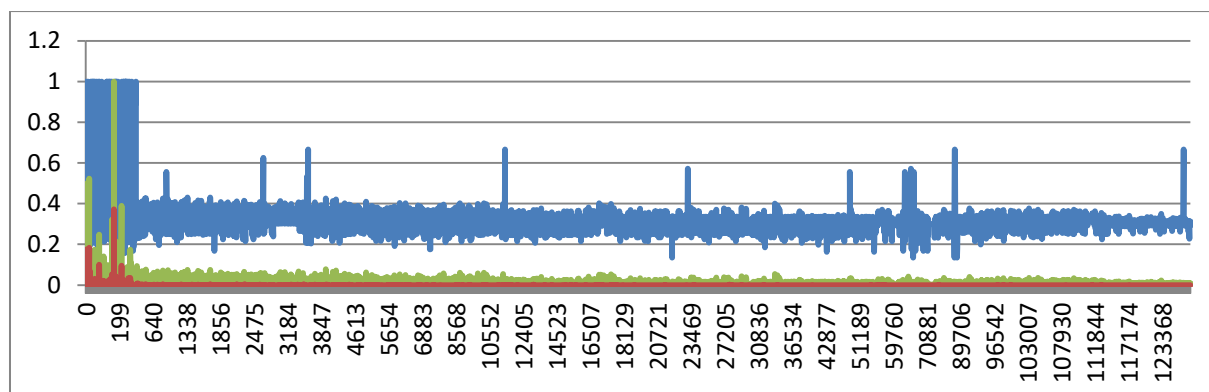
Gigantska komponenta se sastoji od ukupno 4598 čvorova, prosečan koeficijent klasterisanja u njoj je približno 0.17 a postoji više čvorova koji imaju stepen klasterisanja jednak jedan i to su čvorovi koji svojim pozitivnim vezama sa drugim čvorovima grade trouglove.

Ni ovde kao i u prethodnom slučaju neće biti odrađena analiza small-world metrike iz istog razloga kao i tada.

Prosečan stepen klasterisanja na nivou celog grafa iznosi 0.427 a postoji veliki broj čvorovima čiji je

stepen klasterisanja 1. Mrežu odlikuje visok stepen klasterabilnosti. Dijametar mreže je 11. Takođe u analiziranom delu mreže postoji veliki broj visećih čvorova, tj. čvorova koji imaju stepen 1 a njihov stepen klasterisanja je 0, te to obara prosečan stepen klasterisanja.

Kod metrika centralnosti se primećuje da se betweenness i eigenvector centralnost slažu tj. favorizuju iste čvorove i to mahom čvorove sa velikim stepenima u mreži. Međutim closeness centralnost se ne poklapa sa ostale dve metrike. Closeness centralnost favorizuje čvorove sa manjim stepenom, tačnije svi čvorovi koji imaju vrednost ove metrike imaju stepen 1. Dijagram ovih metrika može se videti ispod.



Razlog zašto čvorovi sa manjim stepenom imaju veću closeness centralnost je što su ti čvorovi svojim vezama povezani direktno sa čvorovima koji imaju najveće stepene i koji su povezani sa ostalim čvorovima. Što je zapravo i logično i očekivano jer su neki korisnici ove mreže mnogo aktivniji i samim tim imaju mnogo više oznaka za poverenje ili nepoverenje.

Plava linija predstavlja closeness, zelena eigenvector a crvena betweenness centralnost kao i do sada.

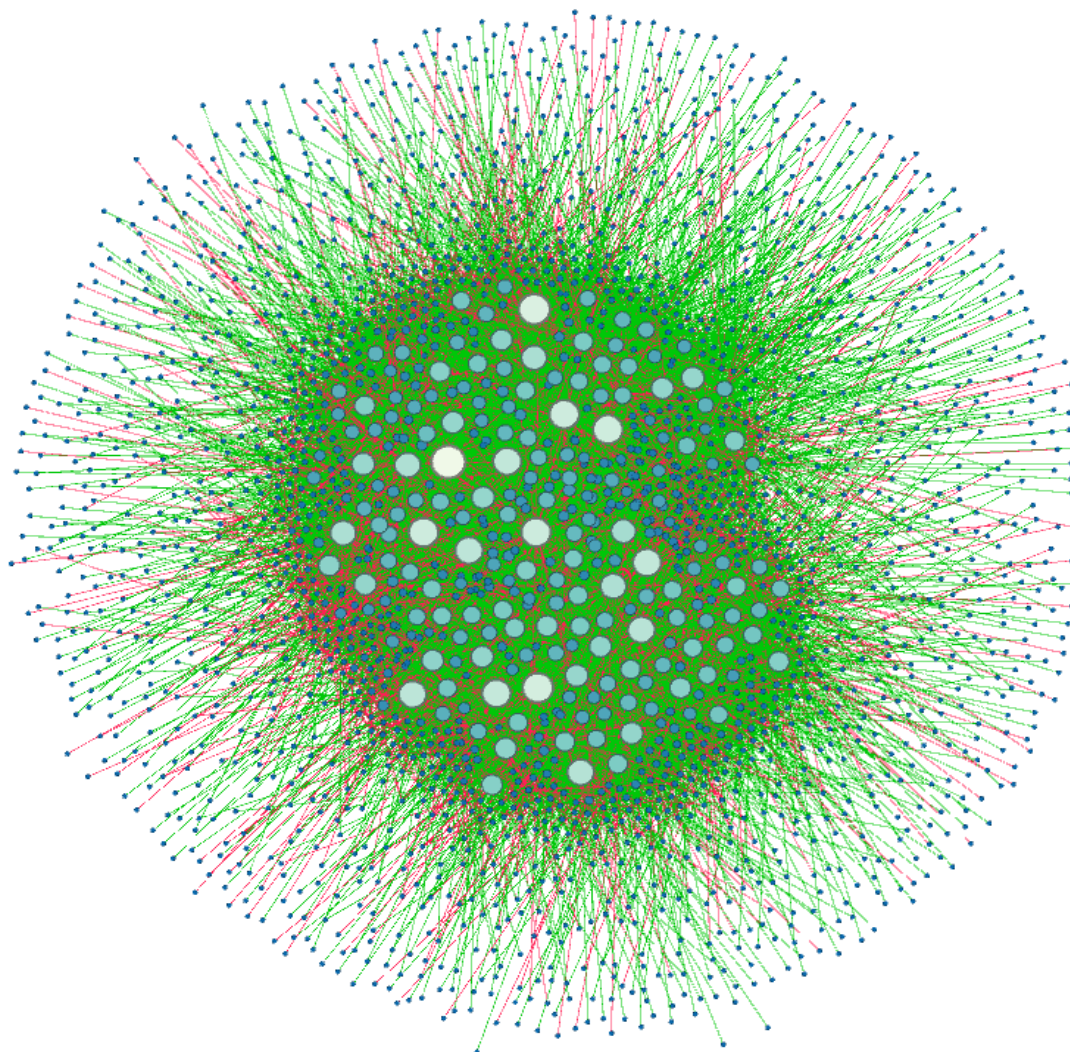
WikiRfA mreža

Da bi član na sajtu Wikipedia postao administrator potrebno je podneti zahtev. Nakon podnošenja zahteva svaki član Wikipedie može podržati kandidaturu biti neutralan ili glasati protiv. Mreža koja će biti analizirana sadrži sve rezultate glasanja od 2003 kada je sistem napravljen pa sve do maja 2013 godine.

Cela mreža se sastoji od 11 381 korisnika i ukupno 198 275 glasova. U fajlu gde se nalazi mreža nalaze se svi glasovi nebitno da li su glasovi podrške, protiv ili neutralni. Prilikom analize mreže neutralni glasovi nisu uračunati tj. preskočeni su.

U uvodnom delu je već rečeno da zbog mogućnosti računara i alata neće biti analizirana cela mreža, međutim u uvodu je rečeno da će biti analizirano prvih 12 000 linija svakog fajla. Ipak kod ove mreže će biti analizirano prvih 20 000 linija zbog same strukture fajla:

```
SRC:Guettarda
TGT:Lord Roem
VOT:1
RES:1
YEA:2013
DAT:19:53, 25 January 2013
TXT:''Support'' per [[WP:DEAL]]: clueful, and unlikely to break
Wikipedia.
```

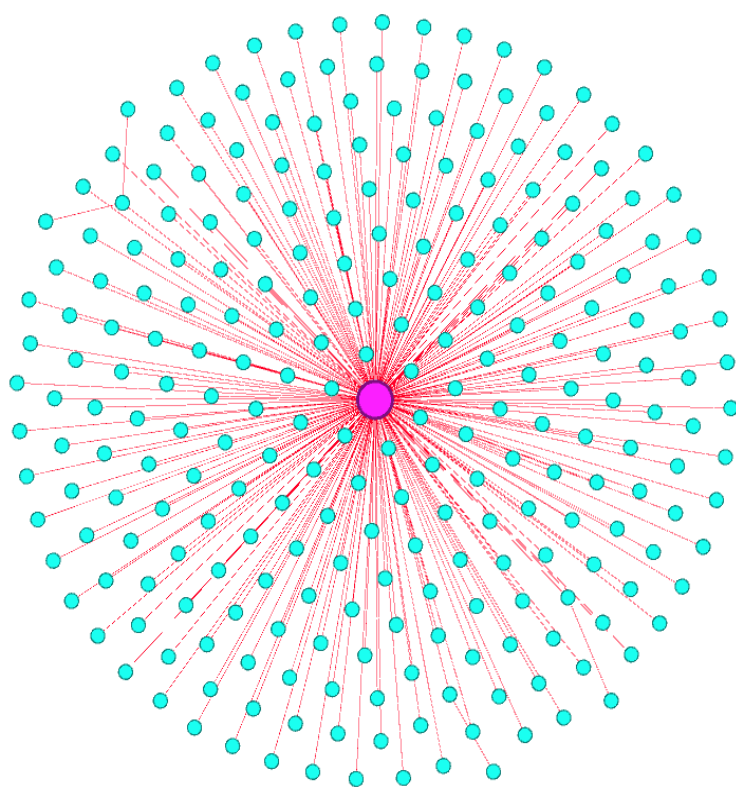
Slika 24

Na slici 24 prikazan je deo mreže koji se analizira. Pravila vizuelizacije su isti kao i kod prethodnih realnih mreža u ovom zadatku.

Klasterabilna	Ukupno klastera	Klasteri koalicije	Klasteri koji nisu koalicije	Broj čvorova	Broj pozitivnih linkova	Broj negativnih linkova
Ne	228	227	1	2265	14191	4064

Mreža nije klasterabilna. Klasterisanjem ulazne mreže dolazi se do 288 klastera od kojih su 227 koalicije a samo 1 nije. Od ukupno 227 koalicija 225 ih se sastoji od samo jednog čvora. Da bi mreža postala klasterabilna potrebno je izbrisati 3580 negativnih linkova od ukupno 4064 koliko ih ima u grafu.

Na slici 25 prikazana je mreža komponenti ulaznog grafa, pravila vizuelizacije su ista kao i kod svih prethodnih primera. Vidi se da je jedini klaster koji nije koalicija upravo gigantska komponenta posmatrane mreže.



Slika 25

Gigantska komponenta se sastoji od 2035 čvorova, ima prosečan koeficijent klasterisanja približno 0.24 i veliki broj čvorova koji imaju stepen klasterisanja 1, to su čvorovi koji svojim vezama obrazuju trouglove čije su strane pozitivne veze.

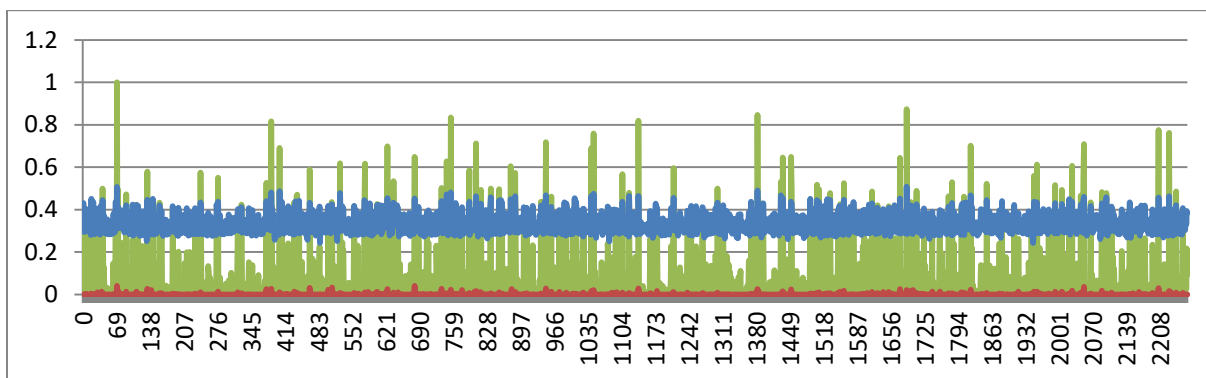
Ni ovde kao i u prethodna dva slučaja neće biti odrađena analiza small-world metrike iz istih razlog kao i do sada

Na nivou cele mreže prosečan stepen klasterisanja iznosi 0.344 a takođe postoji veliki broj čvorovima čiji je stepen klasterisanja 1. Mrežu odlikuje visok stepen klasterabilnosti. Dijametar mreže je 6. Takođe u analiziranom delu mreže postoji veliki broj visećih čvorova, tj. čvorova koji imaju stepen 1 a njihov stepen

klasterisanja je 0, te to obara prosečan stepen klasterisanja a to su mahom viseći čvorovi kojih u ovoj mreži ima velik broj.

Kod metrika centralnosti vidimo se one poklapaju tj. daju prednost istim čvorovima. Takođe metrike centralnosti daju pednost čvorovima sa većim stepenom. Closeness centralnost daje slične vrednosti za sve čvorove, nema velikih razlika u metrici jer su čvorovi uglavnom povezani sa određenim brojem čvorova koji imaju najviše stepene.

Metrike centralnosti prikazane su na grafu ispod, zelenom bojom predstavljena je eigenvector, plavom closeness a crvenom betweenness centralnost.



Zaključak

U ovom dokumentu priložena je implementacija nekih osobina koje su vezane za iznačene mreže a prvenstveno se tiču klasterabilnosti mreže, takođe pored toga odrađena je implementacija četiri modela nasumičnih mreža koja su rađena u okviru kursa i odrađena je analiza tri realne mreže.

U okviru projekta demonstrirano je kako program radi sa malim grafovima koji su osmišljeni tako da znamo šta tačno treba da očekujemo kao izlaz programa.

Zatim je mnogo detaljnije odrađena analiza random mreža. Odrađene su Erdos-Renyi, Gilbert, Watts-Strogatz i Barabasi-Albert modeli redom kako su ovde navođeni. Videli smo neke specifičnosti svake te mreže pojedinačno.

Nakon toga odrađene su analize tri velike realne mreže Slashdot, Epinions i WikiRfA tim redom. Videli smo da se kod realnih mreža mnogo više pitaju korisnici koji su „uticajni“ tj. kod Slashdot mreže mnogo važniji čvorovi i čvorovi koji brže prenose informacije su korisnici sa više prijateljskih odnosa, kod Epinions mreža videli smo da korisnici koji su mnogo aktivniji na mreži imaju mnogo više ocena dok oni drugi uglavnom ocenjuju, a kod Wiki mreže se moglo primetiti da slično kao kod Epinions mreže postoje korisnici koji će mnogo radije da ocene kandidaturu nego neki drugi pa se dobija mreža gde neki korisnici zapravo samo posmatraju interakcije ostalih.

Cela implementacija je odrađena kao proširenje JUNG biblioteke i može se koristiti na bilo kojoj platformi kao sastavni deo iste.

Implementacija se dalje može proširivati za rad sa još nekim vrstama grafova, može se dodavati još modela za kreiranje grafova, dok se implementacija može optimizovati i izmeniti za rad sa većim mrežama, kako bi se u nekom trenutku mogle obraditi cele mreže koje su analizirane u delu o realnim mrežama.